

# On Weakly Bounded Empirical Processes

Shahar Mendelson\*

February 2, 2008

## Abstract

Let  $F$  be a class of functions on a probability space  $(\Omega, \mu)$  and let  $X_1, \dots, X_k$  be independent random variables distributed according to  $\mu$ . We establish high probability tail estimates of the form  $\sup_{f \in F} |\{i : |f(X_i)| \geq t\}|$  using a natural parameter associated with  $F$ . We use this result to analyze weakly bounded empirical processes indexed by  $F$  and processes of the form  $Z_f = \left| k^{-1} \sum_{i=1}^k |f(X_i)| - \mathbb{E}|f|^p \right|$  for  $p > 1$ . We also present some geometric applications of this approach, based on properties of the random operator  $\Gamma = k^{-1/2} \sum_{i=1}^k \langle X_i, \cdot \rangle e_i$ , where the  $(X_i)_{i=1}^k$  are sampled according to an isotropic, log-concave measure on  $\mathbb{R}^n$ .

## 1 Introduction

Empirical Processes theory focuses on understanding the behavior of the supremum of the process

$$f \rightarrow Z_f = \left| \frac{1}{k} \sum_{i=1}^k f(X_i) - \mathbb{E}f \right|$$

where  $F$  is a class of functions on a probability space  $(\Omega, \mu)$ ,  $f \in F$  and  $(X_i)_{i=1}^k$  are independent random variables distributed according to  $\mu$ . Let  $\mu_k$  denote the random empirical measure  $k^{-1} \sum_{i=1}^k \delta_{X_i}$ , and for a class  $F$  we denote the supremum of the empirical process indexed by  $F$  by  $\|\mu_k - \mu\|_F$ . Often, one would like to bound this supremum using geometric properties

---

\*Centre for Mathematics and its Applications, Institute of Advanced Studies, The Australian National University, Canberra, ACT 0200, Australia, and Department of Mathematics, Technion, I.I.T, Haifa 32000, Israel.

email: shahar.mendelson@anu.edu.au

of the set  $F$ , but the question we tackle here is slightly different; our aim is to bound the supremum of the empirical process indexed by powers of the class  $F$ , that is, the supremum of the process indexed by the set  $F^p \equiv \{|f|^p : f \in F\}$  for  $p > 1$  using the geometry of the set  $F$  rather than the geometry of  $F^p$ . The difficulty arises when elements in  $F$  are not necessarily bounded functions, or in cases where the  $L_\infty$  bound is weak - while the situation is considerably simpler in the bounded case. For example, if  $F$  consists of functions bounded by 1 then the empirical process indexed by  $F^p$  can be bounded using a combination of symmetrization and contraction arguments. Indeed, by the Giné-Zinn symmetrization method (see, for example, [5, 25]),

$$\begin{aligned} \mathbb{E}\|\mu_k - \mu\|_{F^p} &\leq 2\mathbb{E} \sup_{f \in F} \left| \frac{1}{k} \sum_{i=1}^k \varepsilon_i |f|^p(X_i) \right| \\ &\leq 2p\mathbb{E} \sup_{f \in F} \left| \frac{1}{k} \sum_{i=1}^k \varepsilon_i f(X_i) \right|, \end{aligned}$$

where  $(\varepsilon_i)_{i=1}^k$  are independent, symmetric  $\{-1, 1\}$ -valued random variables. The last inequality is evident from a contraction principle [9, 25] and the fact that  $|x|^p$  is a Lipschitz function on  $[-1, 1]$  with constant  $p$ .

Moreover, for a class of uniformly bounded functions, the supremum of the empirical process  $\|\mu_k - \mu\|_F$  is highly concentrated around its mean, as the following theorem, due to Talagrand, shows.

**Theorem 1.1** [22, 8] *Let  $F$  be a class of mean zero functions defined on  $(\Omega, \mu)$  such that for every  $f \in F$ ,  $\|f\|_\infty \leq b$ . Let  $X_1, \dots, X_k$  be independent random variables distributed according to  $\mu$  and set  $\sigma^2 = k \sup_{f \in F} \text{var}(f)$ . Define*

$$Z = \sup_{f \in F} \sum_{i=1}^k f(X_i), \quad \bar{Z} = \sup_{f \in F} \left| \sum_{i=1}^k f(X_i) \right|.$$

*Then, for every  $x > 0$ ,*

$$Pr(\{|Z - \mathbb{E}Z| \geq x\}) \leq c_1 \exp\left(-\frac{x}{c_2 b} \log\left(1 + \frac{bx}{\sigma^2 + b\mathbb{E}\bar{Z}}\right)\right), \quad (1.1)$$

*where  $c_1$  and  $c_2$  are absolute constants. The same inequality is also true when  $\bar{Z}$  replaces  $Z$  in (1.1).*

Unfortunately, in many applications the function class at hand does not consist of uniformly bounded functions, or even if the functions are, the

uniform bound is very bad. One such example which motivated this study is the class of linear functionals of Euclidean norm 1 on  $\mathbb{R}^n$ , and the variables  $X_i$  are distributed according to a Borel measure on  $\mathbb{R}^n$  which is natural from the geometric viewpoint, namely, a measure which is isotropic and log-concave.

**Definition 1.2** *A probability measure  $\mu$  on  $\mathbb{R}^n$  is called isotropic if for every  $y \in \mathbb{R}^n$ ,  $\int |\langle x, y \rangle|^2 d\mu(x) = \|y\|^2$ . The measure  $\mu$  is log-concave if for every  $0 < \lambda < 1$  and every Borel measurable  $A, B \subset \mathbb{R}^n$ ,  $\mu(\lambda A + (1 - \lambda)B) \geq \mu(A)^\lambda \mu(B)^{1-\lambda}$ , where  $A + B$  is the Minkowski sum of  $A$  and  $B$ .*

A question of particular interest in this case can be formulated as follows:

**Question 1.3** *Let  $\mu$  be an isotropic measure on  $\mathbb{R}^n$  and let  $X_1, \dots, X_k$  be independent, distributed according to  $\mu$ . Given  $T \subset \mathbb{R}^n$ , for every  $0 < \varepsilon, \delta < 1$  and  $p > 1$ , what is the smallest integer  $k_0$  such that for every  $k \geq k_0$ , with probability at least  $1 - \delta$ ,*

$$\sup_{t \in T} \left| \frac{1}{k} \sum_{i=1}^k |\langle X_i, t \rangle|^p - \mathbb{E} |\langle X, t \rangle|^p \right| < \varepsilon?$$

Two simple examples which come to mind are when  $p = 2$ ,  $T = S^{n-1}$  and  $\mu$  is the Gaussian measure on  $\mathbb{R}^n$  or the uniform measure on the vertices of the unit cube.

**Example 1.4** For every  $t \in \mathbb{R}^n$  define the linear functional  $f_t = \langle t, \cdot \rangle$  and set  $F = \{f_t : t \in S^{n-1}\}$ . Let  $\mu_G$  be the Gaussian measure on  $\mathbb{R}^n$  and note that for every  $t \in S^{n-1}$ ,  $\mathbb{E} f_t^2 = 1$ . Then,

$$\|\mu_k - \mu\|_{F^2} = \sup_{t \in S^{n-1}} \left| \frac{1}{k} \sum_{i=1}^k \langle t, X_i \rangle^2 - 1 \right| = \sup_{t \in S^{n-1}} \left| \frac{1}{k} \|\Gamma t\|^2 - 1 \right|,$$

where  $\Gamma$  is a random  $k \times n$  matrix with independent, standard Gaussian random variables as entries. Hence, if  $\|\mu_k - \mu\|_{F^2} < \varepsilon$ , the gaussian matrix is an almost isometric embedding of  $\ell_2^n$  in  $\ell_2^k$  which is a well known and useful fact and occurs as long as  $k \geq c(\varepsilon, \delta)n$  (see [20]). Another example is when  $\mu = \mu_R$  is the uniform probability measure on  $\{-1, 1\}^n$ . Thus, if  $\|\mu_k - \mu\|_{F^2} \leq \varepsilon$  then a random  $k \times n$  matrix with independent, symmetric,  $\{-1, 1\}$ -valued entries is an almost isometric embedding of  $\ell_2^n$  in  $\ell_2^k$ . Unfortunately, functions in  $F$  on the probability space  $(\mathbb{R}^n, \mu_G)$  are not bounded, while on

$(\mathbb{R}^n, \mu_R)$  the best uniform  $L_\infty$  bound is  $\sup_{t \in S^{n-1}} \|f_t\|_\infty \leq \sqrt{n}$  which is too weak to be useful. Therefore, symmetrization and concentration methods which are so helpful in the bounded case can not assist in resolving Question 1.3 here, as well as in other, more general examples we will explore.

The useful property of linear functionals (with respect to both  $\mu_G$  and  $\mu_R$ ) is that for every  $f_t \in F$ ,

$$\Pr(|f_t| \geq u) \leq 2 \exp(-cu^2)$$

for a suitable absolute constant  $c$ , implying that functions in  $F$  exhibit a subgaussian behavior. Moreover, using Borell's inequality [2, 16], one can show that if  $\mu$  is an arbitrary isotropic log-concave measures, linear functionals exhibit a subexponential decay.

To formulate these decay properties in a more accurate way, we require the definition of Orlicz norms [9, 25].

**Definition 1.5** For  $\alpha \geq 1$  the  $\psi_\alpha$  norm of a random variable  $Y$  is defined by

$$\|Y\|_{\psi_\alpha} = \inf \{u > 0 : \mathbb{E} \exp(|Y|^\alpha/u^\alpha) \leq 2\}.$$

It is standard to verify that if  $Y$  has a bounded  $\psi_\alpha$  norm then  $\Pr(|Y| \geq t) \leq 2 \exp(-ct^\alpha/\|Y\|_{\psi_\alpha}^\alpha)$  where  $c$  is an absolute constant. The reverse direction is also true, and if  $Y$  has a tail bounded by  $\exp(-t^\alpha/K^\alpha)$  then  $\|Y\|_{\psi_\alpha} \leq c_1 K$ .

Our main goal is to show how decay properties of individual class members can be combined to control  $\|\mu_k - \mu\|_{F^p}$ .

As a starting point, let us consider the linear case where is addition, functionals are subgaussian with respect to the  $\ell_2^n$  norm, i.e. for every  $y \in \mathbb{R}^n$ ,  $\|\langle y, X \rangle\|_{\psi_2} \leq c\|y\|_2$ . In particular, the diameter of  $F = S^{n-1}$  is bounded with respect to the  $\psi_2$  norm.

This fact by itself is not enough to bound  $\|\mu_k - \mu\|_{F^p}$ , and to that end we require the following notion of complexity of the class  $F$ .

**Definition 1.6** [24] For a metric space  $(T, d)$ , an admissible sequence of  $T$  is a collection of subsets of  $T$ ,  $\{T_s : s \geq 0\}$ , such that for every  $s \geq 1$ ,  $|T_s| = 2^{2s}$  and  $|T_0| = 1$ . For  $\beta \geq 1$ , define the  $\gamma_\beta$  functional by

$$\gamma_\beta(T, d) = \inf \sup_{t \in T} \sum_{s=0}^{\infty} 2^{s/\beta} d(t, T_s),$$

where the infimum is taken with respect to all admissible sequences of  $T$ .

In [12] the question of estimating  $\|\mu_k - \mu\|_{F^2}$  has been studied for sets of functions which have a bounded diameter with respect to the  $\psi_2$  metric and a finite  $\gamma_2(F, \psi_2)$ , under the additional assumption that for every  $f \in F$ ,  $\mathbb{E}f^2 = 1$ .

**Theorem 1.7** [12] *There exist absolute constants  $c_1, c_2, c_3$  and for which the following holds. Let  $(\Omega, \mu)$  be a probability space, set  $F$  to be a subset of the unit sphere of  $L_2(\mu)$  and assume that  $\text{diam}(F, \psi_2) = \alpha$ . Then, for any  $\theta > 0$  and  $k \geq 1$  satisfying*

$$c_1\alpha\gamma_2(F, \psi_2) \leq \theta\sqrt{k},$$

*with probability at least  $1 - \exp(-c_2\theta^2k/\alpha^4)$ ,  $\|\mu_k - \mu\|_{F^2} \leq \theta$ . Moreover, if  $F$  is symmetric, then  $\mathbb{E}\|\mu_k - \mu\|_{F^2} \leq c_3\alpha\gamma_2(F, \psi_2)/\sqrt{k}$ .*

Theorem 1.7 gives an answer to Question 1.3 for  $p = 2$  under a  $\psi_2$  assumption in a very general situation. It is particularly helpful when the  $\psi_2$  metric endowed on  $F$  is equivalent to the  $L_2$  metric, that is, if for every  $f, g \in F$ ,  $\|f - g\|_{\psi_2} \leq K\|f - g\|_{L_2}$ . In such a case,  $\text{diam}(F, \psi_2) \sim \text{diam}(F, L_2)$  and  $\gamma_2(F, \psi_2) \sim \gamma_2(F, L_2)$ , where by  $A \sim B$  we mean that there are absolute constants  $c$  and  $C$  such that  $cA \leq B \leq CA$ . By the majorizing measures Theorem (see [24] for the most recent survey on the subject),  $\gamma_2(F, L_2)$  is equivalent to the expectation of the supremum of the Gaussian processes indexed by  $F$ , denoted by  $\mathbb{E}\|G\|_F$ . Therefore, under a  $\psi_2$  assumption, Theorem 1.7 implies that if  $F \subset S(L_2)$  then for every  $0 < \delta < 1$ , with probability at least  $1 - \delta$ ,

$$\|\mu_k - \mu\|_{F^2} \leq c \frac{\mathbb{E}\|G\|_F}{\sqrt{k}},$$

where  $c$  depends on  $\delta$  and on the equivalence constant between the  $\psi_2$  and  $L_2$  metrics.

In the geometric context of Example 1.4, Theorem 1.7 is helpful when the indexing set in an arbitrary subset of  $S^{n-1}$ . Moreover, if the measure  $\mu$  happens to be isotropic, then the Gaussian process indexed by  $F$  is the isonormal one and thus  $\gamma_2(F, L_2) \sim \mathbb{E} \sup_{t \in T} |\sum_{i=1}^n g_i t_i|$ , where  $g_1, \dots, g_n$  are independent, standard Gaussian variables.

Unfortunately, the assumption that the  $\psi_2$  metric is equivalent to the  $L_2$  metric is overly optimistic. In particular, the class may not have a well bounded diameter in  $\psi_2$ , or the diameter could be of the same order of magnitude as  $\gamma_2(F, \psi_2)$ . For example, if  $\mu$  is log-concave and isotropic, then for every  $y \in \mathbb{R}^n$ , the function  $f_y = \langle y, \cdot \rangle$  satisfies  $\|f_y\|_{\psi_1(\mu)} \leq K\|f_y\|_{L_2(\mu)}$  and the  $\psi_1$  and  $L_2$  norms are equivalent on  $\mathbb{R}^n$ , but in contrast, the  $\psi_2$

diameter of  $S^{n-1}$  might be polynomial in the dimension (e.g.  $\sqrt{n}$  when  $\mu$  is the normalized volume measure on the isotropic position of the unit ball of  $\ell_1^n$ ). Hence, the bound one can establish from Theorem 1.7 is useless in such cases because of the way it depends on the  $\psi_2$  diameter of the set.

It would be desirable to prove a result of a similar flavor to Theorem 1.7, with the  $\psi_2$  diameter of  $F$  replaced by the  $\psi_1$  diameter and also removes the restrictions that  $p = 2$  and that  $T \subset S(L_2)$ . Our main result implies just that.

To see why the  $\psi_1$  case is considerably more difficult than the  $\psi_2$  one, consider a single function  $h \in L_{\psi_1}$ . By Bernstein's inequality (Lemma 2.2 below), empirical means of  $h$  are highly concentrated around its expectation, with a tail which decays exponentially in sample size. Clearly, if a function  $f \in L_{\psi_2}$  then  $f^2 \in L_{\psi_1}$  and hence exhibits the degree of concentration needed in the proof of Theorem 1.7. On the other hand, if  $f \in L_{\psi_1}$  the degree of concentration of empirical means of  $f^2$  around  $\mathbb{E}f^2$  is not strong enough for that approach.

To overcome this obstacle, the method we suggest here is to decompose  $F$  to two subsets  $F_1$  and  $F_2$  which satisfy that  $F \subset F_1 + F_2$ .

Fix  $\theta(k) > 0$  and consider the sets  $F_1 = \{f 1_{\{|f| \leq \theta\}} : f \in F\}$  and  $F_2 = \{f 1_{\{|f| > \theta\}} : f \in F\}$ .

Since all the functions in  $F_1$  are bounded by  $\theta$ , the empirical mean  $\mu_k(f)$  is highly concentrated around the true mean for any  $f \in F_1$  and  $\|\mu_k - \mu\|_{F_1}$  (or  $\|\mu_k - \mu\|_{F_1^p}$ , using a contraction argument) is well behaved. The key point in this approach is to control the “large part” of the process, namely,

$$\sup_{f \in F} k^{-1} \sum_{i=1}^k |f|^p 1_{\{|f| > \theta\}},$$

and to show that the supremum is small even for a relatively low level of truncation  $\theta$ . The reason this supremum is small has nothing to do with the concentration of each individual class member around its mean, but rather with the fact that with high probability, all the functions in  $F$  have an empirical distribution which decays quickly. And indeed, the main Theorem we present is an “empirical processes” version of result due to Bourgain on the distribution of functions in  $F$  with respect to the (random) empirical measure  $\mu_k$ .

**Theorem A.** *There exist absolute constants  $c_1$ ,  $c_2$  and  $c_3$  for which the following holds. Let  $F$  be a class of mean zero functions on  $(\Omega, \mu)$ . For every  $v_1, v_2 \geq c_1$ , with probability at least  $1 - \exp(-c_2 \min\{v_1, v_2\})$ , for any*

$f \in F$  and  $t > 0$ ,

$$|\{i : |f(X_i)| \geq t\}| \leq \max \left\{ \frac{c_3 v_1 \gamma_2^2(F, \psi_2)}{t^2}, e k \exp \left( -\frac{\theta}{c_3 \alpha v_2} \right) \right\},$$

where  $\alpha = \text{diam}(F, \psi_1)$ .

Bourgain's argument [3] is very different from ours and is tailored to the specific case  $F = \{\langle y, \cdot \rangle : y \in S^{n-1}\}$ , where  $X_1, \dots, X_k$  are selected according to a log-concave measure on  $\mathbb{R}^n$  (see Section 3 for a more detailed discussion).

The proof of Theorem A is based on the following estimate (which will be shown to be optimal) on the  $\ell_1$  structure of a random coordinate projection of  $F$ .

**Theorem B.** *For every  $0 < \delta < 1$  there is a constant  $c(\delta)$  for which the following holds. For every integer  $k$ , with probability at least  $1 - \delta$ , for every  $f \in F$  and  $I \subset \{1, \dots, k\}$ ,*

$$\begin{aligned} \sum_{i \in I} |f(X_i)| &\leq c(\delta) \left( \sqrt{|I|} \gamma_2(F, \psi_2) + \text{diam}(F, \psi_1) |I| \log \left( \frac{ek}{|I|} \right) \right), \\ \sum_{i \in I} |f(X_i)| &\leq c(\delta) \left( \sqrt{|I|} \gamma_2(F, \psi_2) + \text{diam}(F, \psi_2) |I| \sqrt{\log \left( \frac{ek}{|I|} \right)} \right). \end{aligned}$$

We present several geometric applications of Theorem A. The first of which is a “log-concave” version of the celebrated result of Pajor and Tomczak-Jaegermann [17] on sections of small diameter of a convex, symmetric body  $K$  (see also [15, 13, 14] for results along the same lines). We show that if  $X_1, \dots, X_k$  are selected according to an isotropic log-concave measure on  $\mathbb{R}^n$ , then with high probability, the intersection of the kernel of the operator  $\Gamma = \sum_{i=1}^k \langle X_i, \cdot \rangle e_i$  with  $K$  will have a small diameter.

**Theorem C.** *For every  $0 < \delta < 1$  there exists a constant  $c(\delta)$  for which the following holds. Let  $\mu$  be an isotropic, log-concave measure on  $\mathbb{R}^n$  and let  $K \subset \mathbb{R}^n$  be a convex symmetric body. If  $X_1, \dots, X_k$  are independent, distributed according to  $\mu$ , then with probability at least  $1 - \delta$ ,*

$$\text{diam}(\ker \Gamma \cap K) \leq q_k^*(K),$$

where

$$q_k^*(K) = \inf \left\{ \rho > 0 : \rho \geq c(\delta) \frac{V_\rho \sqrt{\log V_\rho}}{\sqrt{k}} \right\},$$

and  $V_\rho = \gamma_2(K \cap \rho S^{n-1}, \psi_2)$ .

If  $\mu$  is a subgaussian measure, Theorem C gives a weaker result (by up to a factor of  $\sqrt{\log n}$ ) and with a weaker probability estimate than Theorem 1.7. On the other hand, it is applicable for a wider set of measures.

The downside of our approach is that it depends on the parameter  $\gamma_2(F, \psi_2)$  which is often hard to bound. However, as we show, a completely  $\psi_1$  version of Theorem B is not true and one might have to use the additional structural assumptions on the indexing set to improve our estimate. Luckily, in the case  $F = \{f_t : t \in S^{n-1}\}$ , it is possible to bound  $\|\mu_k - \mu\|_{F^p}$  in a rather strong sense (though probably suboptimal by a logarithmic factor) using a truncation of the measure  $\mu$ . Let  $\mu$  be a probability measure on  $\mathbb{R}^n$ , for every integer  $k$ , let  $X_1, \dots, X_k$  be independent, distributed according to  $\mu$  and set  $H_k = \mathbb{E} \max_{1 \leq i \leq k} \|X_i\|$ . Observe that if  $Y_i = X_i 1_{\{\|X\| \leq c_1(\delta)H_k\}}$ , then with probability at least  $1 - \delta$ ,  $X_i = Y_i$  for  $1 \leq i \leq k$ . Thus, one can consider the process  $\|\nu_k - \nu\|_{F^p}$  instead of the original process  $\|\mu_k - \mu\|_{F^p}$ . Moreover, one can show

**Theorem D.** *There exist absolute constants  $c_1, c_2$  and  $c_3$  for which the following holds. If  $F = \{f_t : t \in S^{n-1}\}$  then*

$$\gamma_2(F, \psi_2(\nu)) \leq c_1 H_k \sqrt{\log n}.$$

Note that if  $\mu$  is an isotropic log-concave measure on  $\mathbb{R}^n$  and if  $n \leq k \leq \exp(c_2 \sqrt{n})$  then  $H_k \leq c_3 \sqrt{n}$ , which is a fact recently proved by Paouris [19].

As we demonstrate in Section 4, the combination of Theorem B and Theorem D allows us to bound

$$\sup_{t \in S^{n-1}} \left| \frac{1}{k} \sum_{i=1}^k |\langle t, X_i \rangle|^p - \mathbb{E} |\langle t, X \rangle|^p \right|$$

for any log-concave measure.

## 2 Preliminary Results

In this section we present basic results which are used throughout this article. First, a notational convention. All absolute constants are positive numbers, denoted by  $c, c_1, c_2, \dots$  etc. Their value may change from line to line. We denote the Euclidean norm by  $\| \cdot \|$ , while all other norms will be clearly specified.

There are several useful results regarding the concentration and tail behavior of sums of independent random variables. The first one we present

here deals with subgaussian random variables and can be easily seen using the moment generating function.

**Lemma 2.1** [25] *There exists an absolute constant  $c$  for which the following holds. Let  $X$  be a subgaussian random variable and let  $X_1, \dots, X_k$  be independent, distributed as  $X$ . Then, for every  $a = (a_1, \dots, a_k) \in \mathbb{R}^k$*

$$\left\| \sum_{i=1}^k a_i X_i \right\|_{\psi_2} \leq c \|X\|_{\psi_2} \|a\|.$$

If  $X$  is not a  $\psi_2$  random variable and only exhibits a subexponential tail then Bernstein's inequality describes the way the average of independent copies of  $X$  concentrate around their mean - with a tail which is a mixture of subgaussian and subexponential.

**Lemma 2.2** [25] *There exists an absolute constant  $c$  for which the following holds. Let  $X_1, \dots, X_k$  be independent copies of a mean zero random variable. Then, for any  $t > 0$ ,*

$$\Pr \left( \left| \frac{1}{k} \sum_{i=1}^k X_i \right| > t \right) \leq 2 \exp \left( -c k \min \left( \frac{t}{\|X\|_{\psi_1}}, \frac{t^2}{\|X\|_{\psi_1}^2} \right) \right).$$

It turns out that using the generic chaining method [24] combined with Lemma 2.1 or Lemma 2.2, one can bound the supremum of the empirical process indexed by  $F$ .

**Theorem 2.3** [24] *There exists an absolute constant  $c$  for which the following holds. If  $F$  is a class of functions on  $(\Omega, \mu)$ , then for every integer  $k$ ,*

$$\begin{aligned} \mathbb{E} \|\mu_k - \mu\|_F &\leq c \frac{\gamma_2(F, \psi_2)}{\sqrt{k}}, \\ \mathbb{E} \|\mu_k - \mu\|_F &\leq c \left( \frac{\gamma_2(F, \psi_1)}{\sqrt{k}} + \frac{\gamma_1(F, \psi_1)}{k} \right), \end{aligned}$$

and similar bounds hold with high probability.

In many cases, computing the  $\gamma$  functionals is a difficult task. It is possible to upper bound them using a metric entropy integral, similar to Dudley's integral in the context of Gaussian process.

**Definition 2.4** Let  $(T, d)$  be a metric space. The covering number of  $T$  at scale  $\varepsilon$  is the minimal number of open balls (with respect to the metric  $d$ ) of radius  $\varepsilon$  needed to cover  $T$ . The covering numbers of  $(T, d)$  are denoted by  $N(\varepsilon, T, d)$ .

Since one way of forming an admissible sequence for  $(T, d)$  is to use an almost optimal cover (the set  $T_s$  is a cover at the scale at which one needs  $2^{2^s}$  balls to cover  $T$ ), the following is evident:

**Lemma 2.5** There exists an absolute constant  $c$  for which the following holds. Let  $(T, d)$  be a metric space. Then,

$$\gamma_2(T, d) \leq c \int_0^\infty \sqrt{\log N(\varepsilon, T, d)} d\varepsilon.$$

A much more difficult result, due to Talagrand [22, 24], is that if  $T$  is a unit ball of a 2-convex normed space,  $\gamma_2$  could be bounded from above by a sharper version of the entropy integral.

**Definition 2.6** A Banach space is called 2-convex if there is  $\rho > 0$  such that for  $\|x\|, \|y\| \leq 1$ ,  $\|x + y\| \leq 2 - 2\rho\|x - y\|^2$ .

**Theorem 2.7** [23] For every  $\rho > 0$  there exists a constant  $c(\rho)$  for which the following holds. If  $Y$  is a 2-convex Banach space with parameter  $\rho$  and if the metric  $d$  is given by some other norm  $\|\cdot\|$ , then

$$\gamma_2(B_Y, d) \leq c(\rho) \left( \int_0^\infty \varepsilon \log N(B_Y, B_{\|\cdot\|}, \varepsilon) d\varepsilon \right)^{\frac{1}{2}}.$$

Theorem 2.7 is used in the case  $Y = \ell_2^n$ , the  $n$ -dimensional Euclidean space, where  $d$  is the metric endowed on  $\mathbb{R}^n$  by the  $\psi_2$  norm (see Section 4).

### 3 Decomposing classes of functions

Let  $F$  be a class of functions on the probability space  $(\Omega, \mu)$  and assume that for every  $f \in F$ ,  $\mathbb{E}f = 0$ .

Let us formulate the main technical tool we require.

**Theorem 3.1** There exists absolute constants  $c_1$  and  $c_2$  for which the following holds. Let  $F$  be a class of mean zero functions on  $(\Omega, \mu)$  and set  $X_1, \dots, X_k$  to be independent random variables distributed according to  $\mu$ .

Then, for every  $v_1, v_2 \geq c_1$ , with probability at least  $1 - \exp(-c_2 \min\{v_1^2, v_2\})$ , for every  $I \subset \{1, \dots, k\}$ ,

$$\sup_{f \in F} \left| \sum_{i \in I} f(X_i) \right| \leq v_1 \sqrt{|I|} \gamma_2(F, \psi_2) + v_2 \text{diam}(F, \psi_1) |I| \log \left( \frac{ek}{|I|} \right).$$

Theorem 3.1 has a similar version in which one assumes that the set of functions is well bounded in  $\psi_2$ .

**Theorem 3.2** *There exist absolute constants  $c_1$  and  $c_2$  for which the following holds. Let  $F$  and  $X_1, \dots, X_k$  be as in Theorem 3.1. Then, for every  $v \geq c_1$ , with probability at least  $1 - \exp(-c_2 v^2)$ , for every  $I \subset \{1, \dots, k\}$ ,*

$$\sup_{f \in F} \left| \sum_{i \in I} f(X_i) \right| \leq v \left( \sqrt{|I|} \gamma_2(F, \psi_2) + \text{diam}(F, \psi_2) |I| \sqrt{\log \left( \frac{ek}{|I|} \right)} \right).$$

Theorem 3.1 is an empirical processes version of a lemma due to Bourgain ([3], see also [4]) which deals with the case when  $F$  is  $S^{n-1}$ , considered as a class of linear functionals on  $\mathbb{R}^n$  and  $\mu$  is an isotropic log-concave measure. Unlike Bourgain's argument, which relies heavily on the fact that the functions in the class are linear functionals and on that the indexing set is the whole sphere, Theorem 3.1 is very general.

Observe that if the  $L_2$  and  $\psi_2$  metrics are equivalent on  $F$  with a constant  $\beta$  and if  $\mathbb{E}\|G\|_F$  denotes the expectation of the supremum of the Gaussian process indexed by  $F$ , then by the majorizing measures Theorem there are absolute constants  $c$  and  $C$  and a constant  $c_1(\beta)$  depending only on  $\beta$  such that

$$c_1(\beta) \gamma_2(F, \psi_2) \leq c \gamma_2(F, L_2) \leq \mathbb{E}\|G\|_F \leq C \gamma_2(F, L_2) \leq C \gamma_2(F, \psi_2).$$

Therefore, by Theorem 3.1, with probability at least  $1 - \delta$ , for every  $I \subset \{1, \dots, k\}$ ,

$$\begin{aligned} \sup_{f \in F} \left| \sum_{i \in I} f(X_i) \right| &\leq c(\delta, \beta) \left( \sqrt{|I|} \mathbb{E}\|G\|_F + \text{diam}(F, \psi_1) |I| \log \left( \frac{ek}{|I|} \right) \right), \\ \sup_{f \in F} \left| \sum_{i \in I} f(X_i) \right| &\leq c(\delta, \beta) \left( \sqrt{|I|} \mathbb{E}\|G\|_F + \text{diam}(F, \psi_2) |I| \sqrt{\log \left( \frac{ek}{|I|} \right)} \right). \end{aligned}$$

Let us point out that it is impossible to obtain a fully  $\psi_1$  version of Theorem 3.1. Indeed, suppose the converse was true, and that for every set  $F$  and integer  $k$ , with probability at least  $1 - \delta$ , for every  $I \subset \{1, \dots, k\}$ ,

$$\sup_{f \in F} \left| \sum_{i \in I} f(X_i) \right| \leq c(\delta) \left( \sqrt{|I|} \gamma_2(F, \psi_1) + \text{diam}(F, \psi_1) |I| \log \left( \frac{ek}{|I|} \right) \right). \quad (3.1)$$

Let  $Y$  be an exponential random variable and let  $X^{(n)} = \sum_{i=1}^n \frac{Y_i}{\sqrt{\log(i+1)}} e_i \in \mathbb{R}^n$  where  $(e_i)_{i=1}^n$  is the standard basis in  $\mathbb{R}^n$  and  $(Y_i)_{i=1}^n$  are independent copies of  $Y$ . Setting  $\mu^{(n)}$  to be the measure on  $\mathbb{R}^n$  which endows  $X^{(n)}$ , (3.1) can not be true for  $\mu^{(n)}$  and  $F_n = B_1^n$ , the unit ball in  $\ell_1^n$  even when  $k = 1$ . Indeed, using Borell's inequality (see, e.g. [16]) or by a direct computation as in [1], it is evident the for every  $t \in \mathbb{R}^n$ ,

$$\left\| \langle t, X^{(n)} \rangle \right\|_{\psi_1} \leq c \left\| \langle t, X^{(n)} \rangle \right\|_{L_2} = c \left( \sum_{i=1}^n \frac{t_i^2}{\log(i+1)} \right)^{1/2} \equiv c|t|^{(n)},$$

where  $| |^{(n)}$  is the weighted Euclidean norm with weights  $(\sqrt{\log(i+1)})_{i=1}^n$  and  $c$  is an absolute constant. Hence, by the majorizing measures Theorem and a standard computation, there are absolute constants  $c, c_1$  and  $c_2$  such that for every  $n$ ,

$$\gamma_2(F_n, \psi_1(\mu^{(n)})) \leq c \gamma_2(F_n, | |^{(n)}) \leq c_1 \mathbb{E} \sup_{t \in B_1^n} \left| \sum_{i=1}^n g_i \frac{t_i}{\sqrt{\log(i+1)}} \right| \leq c_2.$$

Therefore, if (3.1) were correct for  $k = 1$ , it would follow that with probability of at least  $1/2$ ,  $\sup_{t \in B_1^n} \langle t, X^{(n)} \rangle \leq c_3$ , for a suitable  $c_3$  which is independent of  $n$ . On the other hand, an easy computation shows that with probability larger than some constant  $c_4$ ,

$$\sup_{t \in B_1^n} \langle t, X^{(n)} \rangle \geq \sqrt{\log(n+1)},$$

and thus it is impossible to get a completely  $\psi_1$  version of Theorem 3.1.

Next, observe that Theorem 3.2 is optimal, in the sense that both the  $\gamma_2$  term and the term that depends on the  $\psi_2$ -diameter are required. To see this, fix an integer  $k$  and let  $1 \leq \ell < k$ . Set  $F = \{a, -a\} \subset S^{n-1}$ , acting as linear functional of  $\mathbb{R}^n$  and let  $X = (g_1, \dots, g_n)$  be a Gaussian vector in  $\mathbb{R}^n$ . With this choice of  $X$ , the  $\psi_2$  and  $\ell_2$  metrics on  $\mathbb{R}^n$  are equivalent with an

absolute constant, and thus  $\gamma_2(F, \psi_2) \leq c\gamma_2(F, \ell_2) = c_1$ . On the other hand, writing  $a = (a_1, \dots, a_n)$ , for every  $1 \leq \ell \leq k$ ,

$$\sup_{f \in F} \sup_{I \in E_\ell} \left| \sum_{i \in I} f(X_i) \right| = \sup_{I \in E_\ell} \left| \sum_{i \in I} \sum_{j=1}^n a_j g_{i,j} \right|,$$

which is the supremum of the Gaussian process indexed by

$$E_\ell = \{I : I \subset \{1, \dots, k\}, |I| = \ell\}$$

with the covariance structure endowed by the Hamming metric on  $E_\ell$ , given by  $d_H(I, J) = |I \Delta J|^{1/2}$ . Recall the well known entropy estimate for  $E_\ell$  with respect to this metric (see, for example, [11]):

**Lemma 3.3** *For  $0 < \lambda \leq 1/2$  there exists a constant  $c_\lambda$  for which the following holds. For every integers  $k$  and  $1 \leq \ell \leq k$ , there is a subset  $P \subset E_\ell$  which satisfies that  $\log |P| \geq (1 - \lambda)\ell \log(c_\lambda \frac{k}{\ell})$  and if  $I, J \in P$  and  $I \neq J$  then  $d_H(I, J) \geq \sqrt{\lambda\ell}$ . In other words,*

$$\log N(E_\ell, \sqrt{\lambda\ell}, d_H) \geq (1 - \lambda)\ell \log \left( c_\lambda \frac{k}{\ell} \right).$$

Combining Lemma 3.3 for  $\lambda = 1/4$  with Sudakov's minoration (see, e.g. [9]), it is evident that

$$\mathbb{E} \sup_{I \in E_\ell} \left| \sum_{i \in I} \sum_{j=1}^n a_j g_{i,j} \right| \geq c\ell \sqrt{\log \left( \frac{ck}{\ell} \right)},$$

proving that the second term in Theorem 3.2 is indeed necessary.

To show that the  $\gamma_2$  term is necessary, let  $F = \{-1, 1\}^n$  acting as linear functionals, and again set  $X$  to be the Gaussian vector on  $\mathbb{R}^n$ . Then, for every  $1 \leq \ell \leq k$ ,

$$\sup_{a \in \{-1, 1\}^n} \sup_{I \in E_\ell} \left| \sum_{i \in I} \sum_{j=1}^n g_{i,j} a_j \right| \geq \sup_{a \in \{-1, 1\}^n} \left| \sum_{i=1}^\ell \sum_{j=1}^n g_{i,j} a_j \right|.$$

The latter is the supremum of the Gaussian process indexed by  $\{-1, 1\}^n$  with the covariance structure given by the metric  $d(u, v) = \sqrt{\ell}\|u - v\|_{\ell_2^n}$ . Thus, it is standard to verify that

$$\mathbb{E} \sup_{a \in \{-1, 1\}^n} \left| \sum_{i=1}^\ell \sum_{j=1}^n g_{i,j} a_j \right| \geq c\sqrt{\ell}n.$$

On the other hand,  $\text{diam}(\{-1, 1\}^n, \psi_2) \leq c \text{diam}(\{-1, 1\}^n, \ell_2^n) \leq c\sqrt{n}$ . Thus, the upper bound from Theorem 3.2 is of the order of  $\sqrt{\ell n} + \sqrt{n}\sqrt{\ell \log(ek/\ell)}$ , showing that the  $\gamma_2$  term can not be removed from the bound.

**proof of Theorem 3.1.** To control  $\sup_{f \in F} |\sum_{i \in I} f(X_i)|$ , consider the following  $k$  processes. Recall that for every  $1 \leq \ell \leq k$ ,  $E_\ell = \{I : I \subset \{1, \dots, k\}, |I| = \ell\}$  and define the random process

$$f \rightarrow Z_f^\ell = \sup_{I \in E_\ell} \left| \sum_{i \in I} f(X_i) \right|,$$

where  $X_1, \dots, X_k$  are independent random variables distributed according to  $\mu$ .

Fix  $1 \leq \ell < k$  (the result for  $\ell = k$  requires minor changes and is omitted) and consider the process  $Z_f^\ell$ . Observe that for every  $f, g \in F$ ,

$$\begin{aligned} \Pr \left( |Z_f^\ell - Z_g^\ell| \geq t \right) &\leq \Pr \left( \sup_{I \in E_\ell} \left| \sum_{i \in I} (f - g)(X_i) \right| \geq t \right) \\ &\leq |E_\ell| \Pr \left( \left| \sum_{i=1}^{\ell} (f - g)(X_i) \right| \geq t \right) \\ &\leq 2|E_\ell| \exp \left( -\frac{ct^2}{\|f - g\|_{\psi_2}^2 \ell} \right), \end{aligned}$$

where  $c$  is an absolute constant.

Without loss of generality, assume that  $\gamma_2(F, \psi_2) < \infty$ , let  $(F_s)_{s \geq 0}$  be an almost optimal admissible sequence for the metric space  $(F, \psi_2)$  and set  $\pi_s(f)$  to be a nearest element to  $f$  in  $F_s$  with respect to the  $\psi_2$  metric. Thus,  $|F_s| \leq 2^{2s}$ . Define  $s_0$  as the first index such that  $2^{s_0-1} < \log |E_\ell| \leq 2^{s_0}$ , and note that

$$Z_f^\ell = Z_{\pi_{s_0}(f)}^\ell + \sum_{i=s_0}^{\infty} Z_{\pi_{i+1}(f)}^\ell - Z_{\pi_i(f)}^\ell.$$

Fix  $u > 0$  to be specified later and  $s \geq s_0$ , and consider  $t_s = u\sqrt{\ell} \|\pi_{s+1}(f) - \pi_s(f)\|_{\psi_2} 2^{s/2} \sqrt{\log |E_\ell|}$ . Then,

$$\begin{aligned} \Pr \left( |Z_{\pi_{s+1}(f)}^\ell - Z_{\pi_s(f)}^\ell| \geq t_s \right) &\leq 2|E_\ell| \exp(-cu^2 2^{s-1} \log |E_\ell|) \\ &\leq 2 \exp(-c \log |E_\ell| (u^2 2^{s-1} - 1)). \end{aligned}$$

Take  $u = v_1 / \sqrt{\log |E_\ell|}$  for  $v_1 \geq c_1$  and note that  $2^s > \log |E_\ell|$ , implying that the tail is upper bounded by  $2 \exp(-c_2 v_1^2 2^s)$ . Summing over  $s_0 \leq s < \infty$  it

is evident that with probability at least

$$1 - 2 \sum_{s_0}^{\infty} \exp(-c_2 v_1^2 2^s) \geq 1 - 2 \exp(-c_3 2^{s_0} v_1^2) \geq 1 - 2 \exp(-c_3 v_1^2 \log |E_\ell|),$$

for every  $f \in F$

$$\begin{aligned} \sum_{i=s_0+1}^{\infty} \left| Z_{\pi_{s+1}(f)}^\ell - Z_{\pi_s(f)}^\ell \right| &\leq v_1 \sqrt{\ell} \sum_{i=s_0+1}^{\infty} 2^{s/2} \|\pi_{s+1}(f) - \pi_s(f)\|_{\psi_2} \\ &\leq c_4 v_1 \sqrt{\ell} \gamma_2(F, \psi_2). \end{aligned}$$

To handle  $F_{s_0} = \{\pi_{s_0}(f) : f \in F\}$ , note that the cardinality of this set is at most  $2^{2^{s_0}} \leq 2^{2 \log |E_\ell|}$ . Applying Bernstein's inequality (Lemma 2.2), for every  $t > 0$  and every  $f \in F$ ,

$$\begin{aligned} \Pr \left( \left| Z_f^\ell \right| \geq t\ell \right) &\leq |E_\ell| \Pr \left( \left| \sum_{i=1}^{\ell} f(X_i) \right| \geq t\ell \right) \\ &\leq 2|E_\ell| \exp \left( -c\ell \min \left\{ \frac{t^2}{\|f\|_{\psi_1}^2}, \frac{t}{\|f\|_{\psi_1}} \right\} \right). \end{aligned}$$

Let  $t = \frac{\log |E_\ell|}{\ell} \|f\|_{\psi_1} v_2$  for  $v_2 \geq 1$ . Since  $1 \leq \ell < k$  then  $\ell^{-1} \log |E_\ell| \geq 1$  and  $t \geq \|f\|_{\psi_1}$ . Therefore, with probability at least

$$1 - 2|E_\ell|^2 \exp(-c_5 v_2 \log |E_\ell|) \geq 1 - 2 \exp(-\log |E_\ell| (c_5 v_2 - c_6)),$$

for every  $f \in F_{s_0}$ ,

$$Z_f^\ell \leq v_2 \|f\|_{\psi_1} \log |E_\ell| \leq v_2 \text{diam}(F, \psi_1) \log |E_\ell|.$$

To conclude, there are absolute constants  $c_7, c_8$  and  $c_9$  such that for every  $1 \leq \ell \leq k$ , if  $v_1, v_2 \geq c_7$ , with probability at least  $1 - 2 \exp(-c_8 \log |E_\ell| \min\{v_1^2, v_2\})$ ,

$$\sup_{f \in F} |Z_f^\ell| \leq c_9 \left( v_1 \sqrt{\ell} \gamma_2(F, \psi_2) + v_2 \text{diam}(F, \psi_1) \log |E_\ell| \right).$$

Summing the probabilities, the latter holds for every  $1 \leq \ell \leq k$  with probability at least  $1 - \exp(-c_{10} \min\{v_1^2, v_2\})$ , completing the proof.  $\blacksquare$

The proof of Theorem 3.2 is similar and is omitted.

**Proof of Theorem B.** In the  $\psi_1$  case, take  $v_1 = \sqrt{\log(1/\delta)}$  and  $v_2 = \log(1/\delta)$  for  $\delta$  small enough. Fix any  $f \in F$  and for  $I \in E_\ell$  let  $I^+(f) = \{i :$

$f(X_i) > 0\} \cap I$  and  $I^-(f) = \{i : f(X_i) < 0\} \cap I$ . Then, by Theorem 3.1, with probability at least  $1 - \delta$ ,

$$\begin{aligned} \sum_{i \in I} |f(X_i)| &= \left| \sum_{i \in I^+(f)} f(X_i) \right| + \left| \sum_{i \in I^-(f)} f(X_i) \right| \\ &\leq c(\delta) \left( \sqrt{\ell} \gamma_2(F, \psi_2) + \text{diam}(F, \psi_1) \ell \log \left( \frac{ek}{\ell} \right) \right), \end{aligned}$$

as claimed. The  $\psi_2$  case is equally easy.  $\blacksquare$

For Theorem 3.1 one can derive the following uniform empirical tail estimate for functions in  $F$ , which was formulated as Theorem A in the introduction.

**Corollary 3.4** *There exist absolute constants  $c_1, c_2$  and  $c_3$  for which the following holds. Let  $F$  be as in Theorem 3.1. For every  $v_1, v_2 \geq c_1$ , with probability at least  $1 - \exp(-c_2 \min\{v_1^2, v_2\})$ , for any  $f \in F$  and  $t > 0$ ,*

$$|\{i : |f(X_i)| \geq t\}| \leq \max \left\{ \frac{c_3 v_1^2 \gamma_2^2(F, \psi_2)}{t^2}, ek \exp \left( -\frac{t}{c_3 \alpha v_2} \right) \right\}, \quad (3.2)$$

where  $\alpha = \text{diam}(F, \psi_1)$ .

**Proof.** Fix  $v_1, v_2$  as in Theorem 3.1 and consider the set for which the assertion of Theorem 3.1 holds. Let  $(X_1, \dots, X_k)$  be in that set and for  $f \in F$  and  $t > 0$  put

$$I_t(f) = \{i : |f(X_i)| \geq t\}.$$

Setting  $\alpha = \text{diam}(F, \psi_1)$  there are two possibilities. First, if

$$2\alpha v_2 |I_t(f)| \log \left( \frac{ek}{|I_t(f)|} \right) \leq \frac{t}{2} |I_t(f)|,$$

then by Theorem 3.1,

$$\begin{aligned} t |I_t(f)| &\leq 2v_1 \sqrt{|I_t(f)|} \gamma_2(F, \psi_2) + 2v_2 \alpha |I_t(f)| \log \left( \frac{ek}{|I_t(f)|} \right) \\ &\leq 2v_1 \sqrt{|I_t(f)|} \gamma_2(F, \psi_2) + \frac{t}{2} |I_t(f)|. \end{aligned}$$

Thus,  $t |I_t(f)| / 2 \leq 2v_1 \sqrt{|I_t(f)|} \gamma_2(F, \psi_2)$ , implying that

$$|I_t(f)| \leq 16v_1^2 \frac{\gamma_2^2(F, \psi_2)}{t^2}.$$

Otherwise,  $2\alpha v_2 |I_t(f)| \log(ek/|I_t(f)|) \geq t|I_t(f)|/2$ , or in other words,

$$|I_t(f)| \leq ek \exp\left(-\frac{t}{4v_2 \text{diam}(F, \psi_1)}\right).$$

■

Now we are ready to formulate and prove the main theorem of this section, which is a decomposition result for the class  $F$ .

**Theorem 3.5** *There exist absolute constants  $c_1, c_2$  and  $c_3$ , and for every  $1 \leq p < \infty$  there exists a constants  $c_4(p)$  for which the following holds. Let  $F$  be a class of mean zero functions. For  $v \geq c_1$ ,  $A \geq \gamma_2(F, \psi_2)$ ,  $B \geq \text{diam}(F, \psi_1)$  and an integer  $k$ , set*

$$\theta \geq \max \left\{ c_2 v B \log \left( c_2 \frac{B^2 k}{A^2} v + 1 \right), c_2 p B \log(c_2 p B + 1) \right\}.$$

*Then, there are Lipschitz functions  $\phi : \mathbb{R} \rightarrow \mathbb{R}$  and  $\psi : \mathbb{R} \rightarrow \mathbb{R}$  which depend on  $\theta$ , such that  $\|\phi\|_{\text{lip}} = \|\psi\|_{\text{lip}} \leq 1$  and setting  $F_1 = \{\phi(f) : f \in F\}$  and  $F_2 = \{\psi(f) : f \in F\}$ ,*

1.  $F \subset F_1 + F_2$ .
2. For every  $h \in F_1$ ,  $\|h\|_\infty \leq \theta$  and for every  $h \in F_2$ ,  $\mathbb{E}|h|^p \leq A^2/k$ .
3. With probability at least  $1 - \exp(-c_3 v)$ ,

$$\sup_{h \in F_2} \sum_{i=1}^k |h(X_i)|^p \leq c_2 v A^2 (\theta^{p-2} + \kappa_p),$$

where  $\kappa_p = c_4(p) \theta^{p-2}$  for  $p < 2$ ,  $\kappa_2 = c_4(2) \log A$ , while for  $p > 2$ ,  $\kappa_p = c_4(p) A^{p-2}$ .

Theorem 3.5 implies that  $F$  can be decomposed into two simple sets  $F_1$  and  $F_2$  (which depend on  $k, p$  and  $v$ ). The fact that these sets are as simple as  $F$  is evident because they are images of  $F$  via Lipschitz functions with constant 1. In particular,  $\gamma_\beta(F_i, d) \leq \gamma_\beta(F, d)$  with respect to any reasonable metric  $d$ . The sets  $F_1$  and  $F_2$  have additional properties.  $F_1$  has a bounded diameter in  $L_\infty$  - up to a logarithmic term, its diameter in  $L_\infty$  is proportional to the  $\psi_1$  diameter of  $F$ . Thus, if  $F$  has a well bounded diameter with respect to the  $\psi_1$  metric then functions in  $F_1$  are highly concentrated around their means, and one can safely use a contraction

argument when bounding the empirical process indexed by a power of  $F_1$ . The main difficulty is in controlling the “large part” of  $F$  - i.e.  $F_2$ . The empirical process indexed by  $F_2^p$  is small not because of concentration, but because the  $\ell_p^k$  diameter of a random coordinate projection of  $F_2$  and its  $L_p$  diameter are small.

**Proof of Theorem 3.5.** Fix an integer  $k$  and  $v$  for which Corollary 3.4 holds. The first step is to select the Lipschitz functions  $\phi$  and  $\psi$ ; those are simply truncation functions at the level  $\theta$ . For  $f \in F$ , set  $\phi(f) = \text{sgn}(f) \min\{|f|, \theta\}$  and  $\psi(f) = f - \phi(f)$ . Clearly, both functions have Lipschitz constant 1,  $F \subset \phi(F) + \psi(F)$ , and for  $p > 1$ , because  $\phi(f)$  and  $\psi(f)$  are supported on disjoint sets,

$$\begin{aligned} |f|^p &= \min\{|f|^p, \theta^p\} + (|f|^p - \theta^p) 1_{\{|f| \geq \theta\}} \\ &\leq |\phi(f)|^p + |f|^p 1_{\{|f| \geq \theta\}}. \end{aligned}$$

Let  $A \geq \gamma_2(F, \psi_2)$  and  $B \geq \text{diam}(F, \psi_1)$ . It is evident that

$$\mathbb{E}|f|^p 1_{\{|f| \geq \theta\}} \leq c_1(2pB)^p \exp\left(-\frac{\theta}{c_2 B}\right) \leq \frac{A^2}{k}, \quad (3.3)$$

as long as

$$\theta \geq (c_3 p B) \log(c_3 p B) + c_3 B \log\left(\frac{B^2 k}{A^2}\right), \quad (3.4)$$

which is satisfied by our choice of  $\theta$ . Thus (2) is established.

Turning to (3), recall that for every  $t > 0$  and  $f \in F$ ,  $I_t(f) = \{i : |f(X_i)| \geq t\}$ . By our choice of  $v$ , with probability at least  $1 - \exp(-c_4 v)$ , for every  $f \in F$ , for every  $t > 0$

$$|I_t(f)| \leq \max\left\{\frac{c_5 v A^2}{t^2}, k \exp\left(-\frac{t}{c_5 B v}\right)\right\}. \quad (3.5)$$

Therefore, if

$$t \geq t_0 = c_6 B v \log(c_6 B v \sqrt{k}/A) \quad (3.6)$$

for a suitable absolute constant  $c_6$ , the first term in (3.5) is dominant. Note that if  $t \geq \max\{t_0, \sqrt{c_5 v} A\}$  then  $|I_t(f)| = 0$ , and since  $\theta \geq t_0$  then by a standard integration argument with respect to the random empirical measure  $\mu_k$ , with probability at least  $1 - \exp(-c_4 v)$ , for every  $f \in F$

$$\begin{aligned} \mathbb{E}_{\mu_k} |f|^p 1_{\{|f| \geq \theta\}} &\leq \theta^p \Pr_{\mu_k} (|f| \geq \theta) + \int_{\theta}^{\sqrt{c_5 v} A} p t^{p-1} \Pr_{\mu_k} (|f| > t) dt \\ &\leq c_6 \frac{v A^2}{k} \left( \theta^{p-2} + \int_{\theta}^{\sqrt{c_5 v} A} p t^{p-3} dt \right), \end{aligned}$$

for which the claim follows.  $\blacksquare$

**Remark 3.6** Note that Theorem 3.5 enables one to bound  $\|\mu_k - \mu\|_{F_1}$  and thus  $\|\mu_k - \mu\|_F$ . Indeed, pointwise,  $|(\phi(f))^p - (\phi(g))^p| \leq 2p\theta^{p-1}|f - g|$ , implying that

$$\gamma_2((\phi(F))^p, \psi_2) \leq cp\theta^{p-1}\gamma_2(F, \psi_2).$$

By a standard generic chaining argument (see Theorem 2.3 and [24]), for every  $v > 0$ , with probability at least  $1 - \exp(-c_1v^2)$ ,

$$\begin{aligned} \sup_{f \in F} \left| \frac{1}{k} \sum_{i=1}^k (\phi(f))^p (X_i) - \mathbb{E} (\phi(f))^p \right| &\leq c_2 v \frac{\gamma_2((\phi(F))^p, \psi_2)}{\sqrt{k}} \\ &\leq c_3 p \theta^{p-1} v \frac{\gamma_2(F, \psi_2)}{\sqrt{k}}. \end{aligned}$$

Combining this with Theorem 3.5, it follows that with probability at least  $1 - 2\exp(-c_1v)$ ,

$$\|\mu_k - \mu\|_F \leq c_2 v \left( p\theta^{p-1} \frac{\gamma_2(F, \psi_2)}{\sqrt{k}} + \frac{\gamma_2^2(F, \psi_2)}{k} (\theta^{p-2} + \kappa_p + 1) \right).$$

**Remark 3.7** Observe that by (3.3),  $\sup_{h \in F_2} \mathbb{E}|h|^p \leq (cpB)^p \exp(-\frac{\theta}{cB})$ , a fact we shall use below.

We end this section with another observation which follows easily from the proof of Theorem 3.5. To avoid complications, we will formulate it only in the case we need it, which is when  $F$  is a class of linear functionals on  $\mathbb{R}^n$  and  $\mu$  is a measure on  $\mathbb{R}^n$ . Consider the random variable  $U = \sup_{f \in F} |\langle f, X \rangle|$ , and for every integer  $k$  set  $H_k = \mathbb{E} \max_{1 \leq i \leq k} U_i$ , where  $(U_i)_{i=1}^k$  are independent copies of  $U$ .

**Theorem 3.8** For every  $p \geq 1$  and  $0 < \delta, \varepsilon < 1$  there are constants  $c_1(\delta, \varepsilon, p)$ ,  $c_2(\delta, p)$  and  $c_3(p)$  for which the following holds. Let  $F$  and  $\mu$  be as above, consider the random variable  $Y^k = X \mathbb{1}_{\{U \leq c_1(\delta, \varepsilon, p)H_k\}}$  and let  $\nu$  be the probability measure on  $\mathbb{R}^n$  corresponding to  $Y^k$ . If  $A \geq \gamma_2(F, \psi_2(\nu))$  and  $B \geq \text{diam}(F, \psi_1(\nu))$ , then with probability at least  $1 - \delta$ ,

$$\|\mu_k - \mu\|_{F^p} \leq c_2 \left( \theta^{p-1} \frac{\gamma_2(F, \nu)}{\sqrt{k}} + \frac{\gamma_2^2(F, \nu)}{k} (\theta^{p-2} + \tilde{\kappa}_p) \right) + c_3 B^{1/2} \varepsilon, \quad (3.7)$$

where

$$\theta = \max \left\{ c_2 B \log \left( c_2 \frac{kB^2}{A^2} + 1 \right), c_2 p B \log(c_2 p B + 1) \right\},$$

and  $\tilde{\kappa}_p = 1$  for  $1 \leq p < 2$ ,  $\tilde{\kappa}_2 = \log H_k$  and  $\tilde{\kappa}_p = H_k^{p-2}$  for  $p > 2$ .

Because the proof is based on the same arguments used in Theorem 3.5 and Remark 3.6, we will only give a brief sketch of the required modifications which are that with high probability,  $X_i = Y_i$  for  $1 \leq i \leq k$  and that by the Cauchy-Schwarz inequality,

$$\sup_{f \in F} |\mathbb{E}_\mu |f|^p - \mathbb{E}_\nu |f|^p| = \sup_{f \in F} \mathbb{E} |f|^p 1_{\{U \geq c_1(\delta, \varepsilon, p) H_k\}} \leq c_3(p) B^{1/2} \varepsilon,$$

for the right choice of constants. Thus, one can replace the measure  $\mu$  with the measure  $\nu$  and consider the empirical process  $\|\nu - \nu_k\|_{F^p}$  instead of  $\|\mu - \mu_k\|_{F^p}$ .

The advantage of using the measure  $\nu$  is that it is a truncated version of  $\mu$  at the “correct” level for  $F$  and the sample size  $k$ . This truncation enables us to bound  $\gamma_2(S^{n-1}, \nu)$ , where  $\nu$  is a truncation of an isotropic, log-concave measure on  $\mathbb{R}^n$ .

## 4 Applications

The first geometric application we present deals with sections of small diameter of a convex, symmetric body. Let  $\Gamma = \sum_{i=1}^k \langle X_i, \cdot \rangle e_i$ , where  $X_1, \dots, X_k$  are selected according an isotropic log concave measure. As we show below, if  $K$  is a convex, symmetric body in  $\mathbb{R}^n$ , then with high probability, the diameter of  $K \cap \ker(\Gamma)$  is small. This extends a celebrated result of Pajor and Tomczak-Jaegermann [17] which was proved in the case where the random subspace was selected according to the Haar measure on the Grassmann manifold  $\mathcal{G}(n, k)$ , but the same proof works in the Gaussian case. Various versions and extensions of this result may be found, for example, [15, 13, 14].

The following theorem is a formulation of version of this result for a general  $\psi_2$  operator (see [12]). Let us introduce the following notation: for a set  $T \subset \mathbb{R}^n$  we denote by  $\ell_*(T) = \mathbb{E} \sup_{t \in T} |\sum_{i=1}^n g_i t_i|$ , the expectation of the supremum of the gaussian process indexed by  $T$ . Recall that by the majorizing measures Theorem [24], there are absolute constants  $c_1$  and  $c_2$  such that for every  $T \subset \mathbb{R}^n$ ,

$$c_1 \gamma_2(T, \|\cdot\|) \leq \ell_*(T) \leq c_2 \gamma_2(T, \|\cdot\|). \quad (4.1)$$

**Theorem 4.1** [12] *There exists a absolute constant  $c$  and  $c_1$  for which the following holds. Let  $X_1, \dots, X_k$  be distributed according to an isotropic measure  $\mu$  on  $\mathbb{R}^n$  and assume that for every  $t \in \mathbb{R}^n$ ,  $\|\langle t, \cdot \rangle\|_{\psi_2} \leq \alpha \|t\|$  for some  $\alpha \geq 1$ . If  $K \subset \mathbb{R}^n$  is a convex symmetric body then with probability at least  $1 - \exp(-c_1 k / \alpha^4)$ ,*

$$\text{diam}(\ker \Gamma \cap K) \leq r_k^*(K),$$

where

$$r_k^*(K) = \inf \left\{ \rho > 0 : \rho \geq c_1 \alpha^2 \ell_*(K \cap \rho S^{n-1}) / \sqrt{k} \right\}.$$

Our result is similar (though with a weaker estimate) to Theorem 4.1. Other than the different ways of estimating the empirical process  $\|\mu_k - \mu\|_{F^2}$ , the two proofs are identical, and thus the proof of Theorem 4.2 is omitted.

**Theorem 4.2** *For every  $0 < \delta < 1$  there exist constant  $c(\delta)$  for which the following holds. Let  $\mu$  be an isotropic, log-concave measure on  $\mathbb{R}^n$  and let  $K \subset \mathbb{R}^n$  be a convex symmetric body. If  $X_1, \dots, X_k$  are independent, distributed according to  $\mu$ , then with probability at least  $1 - \delta$ ,*

$$\text{diam}(\ker \Gamma \cap K) \leq q_k^*(K),$$

where

$$q_k^*(K) = \inf \left\{ \rho > 0 : \rho \geq c(\delta) \gamma_2(K \cap \rho S^{n-1}, \psi_2) \sqrt{\frac{\gamma_2(K \cap \rho S^{n-1}, \psi_2)}{k}} \right\}.$$

If  $\mu$  is a subgaussian measure then for every  $A \subset \mathbb{R}^n$ ,  $\gamma_2(A, \psi_2) \leq c \ell_*(A)$ , and thus  $\gamma_2(K \cap \rho S^{n-1}, \psi_2) \leq \sqrt{n}$ . Therefore, Theorem 4.2 recovers Theorem 4.1 up to a  $\sqrt{\log n}$  factor. Of course, the bound on the probability is considerably weaker. On the other hand, Theorem 4.2 holds for a much wider family of measures because the bound given in Theorem 4.1 depends on the equivalence constant between the  $\psi_2$  and  $\ell_2$  metrics endowed on  $\mathbb{R}^n$ .

#### 4.1 Sampling from an isotropic, log-concave measure

A question which was originally studied in [7, 3, 21, 4, 6] is the following: how many points sampled from an isotropic, convex, symmetric body are needed to ensure that the random operator  $k^{-1} \sum_{i=1}^k \langle X_i, \cdot \rangle e_i$  is an almost isometric embedding of  $\ell_2^n$  in  $\ell_2^k$ ? In other words, that with probability at least  $1 - \delta$ , for every  $\theta \in S^{n-1}$ ,

$$1 - \varepsilon \leq \frac{1}{k} \sum_{i=1}^k \langle X_i, \theta \rangle^2 \leq 1 + \varepsilon.$$

**Theorem 4.3** [21, 4, 6] For every  $0 < \varepsilon, \delta < 1$  there is a constant  $c(\varepsilon, \delta)$  for which the following holds. Let  $X_1, \dots, X_k$  be independent random variables, distributed according to the volume measure of a convex, symmetric body in isotropic position. If  $k \geq c(\varepsilon, \delta)n \log^2 n$ , then with probability at least  $1 - \delta$ , for every  $\theta \in S^{n-1}$ ,

$$1 - \varepsilon \leq \frac{1}{k} \sum_{i=1}^k \langle \theta, X_i \rangle^2 \leq 1 + \varepsilon.$$

The estimate of  $k \sim n \log^2 n$  was first proved by Rudelson [21]. Previously, Bourgain showed [3] how to obtain this result with a slightly weaker estimate of  $k \sim n \log^3 n$ , but then Giannopoulos and Milman [4] demonstrated that Bourgain's method can actually give the same estimate as Rudelson's.

The proofs of Bourgain and Rudelson use very different arguments. Rudelson's proof is based on a noncommutative Khintchine inequality, due to Lust-Piquard and Pisier [10], namely, a bound on Rademacher averages of the form  $\mathbb{E} \left\| \sum_{i=1}^k \varepsilon_i x_i \otimes x_i \right\|_{\ell_2 \rightarrow \ell_2}^p$  for  $p \geq 1$ , where  $(x_i)_{i=1}^k \in \ell_2$  and  $(\varepsilon_i)_{i=1}^k$  are independent, symmetric,  $\{-1, 1\}$ -valued random variables. The fact that the set indexing the empirical process is exactly the Euclidean sphere is essential in the proof and the argument can not be modified to handle any other indexing sets - not even other subsets of the sphere.

Bourgain's proof uses a similar technique to the one we used here, which relies on the following version of Theorem 3.1. The formulation we present here is from [4].

**Lemma 4.4** Let  $\delta \in (0, 1)$  and let  $X_1, \dots, X_k$  be points in  $\mathbb{R}^n$  sampled according to an isotropic log-concave measure. If  $k \leq c\delta \exp(\sqrt{n})$  then with probability at least  $1 - \delta$ , for every  $I \subset \{1, \dots, k\}$ ,

$$\left\| \sum_{i \in I} X_i \right\| \leq c_1(\delta) \left( \sqrt{\log k} \sqrt{|I|} \sqrt{n} + |I| \log k \right).$$

In particular, with probability at least  $1 - \delta$ , for every  $t \geq c(\delta) \log k$  and every  $x \in S^{n-1}$ ,

$$|\{i : \langle x, X_i \rangle \geq t\}| \leq c_2(\delta) \frac{n \log k}{t^2}. \quad (4.2)$$

Bourgain's method was generalized in [4] in which the following theorem was established:

**Theorem 4.5** *Let  $p > 0$  and  $0 < \delta < 1$ . There exists  $n_0(\delta)$  such that for every  $n \geq n_0(\delta)$ , every log-concave measure on  $\mathbb{R}^n$ , every  $k \geq k_0(\delta, p)$  and every  $\theta \in S^{n-1}$ ,*

$$c_p \leq \frac{1}{k} \sum_{i=1}^k |\langle X_i, \theta \rangle|^p \leq C_p$$

where  $c_p$  and  $C_p$  depend only on  $p$  and

$$k_0(\delta, p) = c(\delta, p) \begin{cases} n & \text{if } 0 < p < 1, \\ n \log^p n & \text{if } 1 \leq p \leq 2, \\ \min\{(p-2)^{-1}, \log n\}(n \log n)^{p/2} & \text{if } p > 2. \end{cases}$$

Note that this bound is isomorphic in nature rather than almost isometric, though in the case  $p = 2$  the proof of Theorem 4.5 can be modified to give an almost isometric estimate.

Recently, Guédon and Rudelson [6] were able to bound

$$\mathbb{E}_\varepsilon \sup_{y \in K} \sum_{i=1}^k \varepsilon_i |\langle x_i, y \rangle|^p,$$

for any  $x_1, \dots, x_k \in \mathbb{R}^n$ , where  $K \subset B_2^n$  is a convex, symmetric body which has a  $q$ -power type modulus of convexity. The method of proof is based on majorizing measures, and can be used to bound  $\mathbb{E}\|\mu_k - \mu\|_{F^p}$  for  $F = \{\langle x, - \rangle : x \in K\}$  as long as  $p \geq q \geq 2$ . It turns out that the dominant factor in the bound is

$$\left( \mathbb{E} \max_{1 \leq i \leq k} \|X_i\|^2 \cdot \mathbb{E} \max_{1 \leq i \leq k} \|X_i\|_{K^\circ}^{p-2} \right)^{\frac{1}{p}}.$$

For  $K = B_2^n$  this approach yields the best known estimates for  $\mathbb{E}\|\mu_k - \mu\|_{F^p}$  for  $p \geq 2$ , and the resulting estimate on the required size of the sample is  $k \sim c(\varepsilon, \delta, p)n^{p/2} \log n$ , and in particular, for  $p = 2$  gives the best known estimate of  $k \sim c(\varepsilon, \delta)n \log n$ . Let us mention that for  $p = 2$  this result is not helpful for “small” subsets of the sphere, and the best bound that one can establish for such subsets coincides with the one obtained for the whole sphere.

All the known bounds, including [6] and ours, are based on the behavior of the random variable  $\|X\|$ . The best estimates on  $\|X\|$  are due to Paouris [19]:

**Theorem 4.6** *There are absolute constants  $c_1$  and  $c_2$  for which the following holds. Let  $X$  be distributed according to an isotropic log-concave measure on  $\mathbb{R}^n$ . Then, for every  $p \leq c_1\sqrt{n}$ ,  $(\mathbb{E}\|X\|^p)^{1/p} \leq c_2\sqrt{n}$ .*

Theorem 4.6 immediately leads to a removal of a logarithmic factor in (4.2), though not to an improved level of truncation; thus, the estimate of Theorem 4.5 in the case  $p = 2$  remains unchanged despite the improved tail estimate.

The properties of an isotropic log-concave measure which will be used below are that for suitable absolute constants  $C$  and  $C_1$ ,

1. linear functionals have a subexponential tail - that is, for every  $x \in \mathbb{R}^n$ ,

$$\|\langle X, \cdot \rangle\|_{\psi_1} \leq C\|x\|,$$

and

2. By Theorem 4.6, for  $n \leq k \leq \exp(c\sqrt{n})$ ,

$$\mathbb{E} \max_{1 \leq i \leq k} \|X_i\| = \mathbb{E} \max_{1 \leq i \leq k} \sup_{x \in S^{n-1}} |\langle x, X_i \rangle| \leq c_1\sqrt{n}.$$

Therefore, in light of Theorem 3.8, it is enough to consider the truncated measure  $\nu$  on  $\mathbb{R}^n$  which is supported on a ball of radius  $c_2(\delta)\sqrt{n}$  to bound  $\|\mu_k - \mu\|_{F^p}$ . The main ingredient in our method is to bound  $\gamma_2(S^{n-1}, \nu)$  and to that end we shall estimate

$$\ell_E = \mathbb{E} \left\| \sum_{i=1}^n g_i e_i \right\|_E,$$

where  $g_1, \dots, g_n$  are standard, independent Gaussian variables. The particular norm  $\|\cdot\|_E$  we consider is the one endowed on  $\mathbb{R}^n$  by the  $\psi_2(\nu)$  structure, formally defined for every  $t \in \mathbb{R}^n$  by  $\|t\|_E = \|\langle t, Y \rangle\|_{\psi_2}$ .

**Lemma 4.7** *There exists an absolute constant  $c$  for which the following holds. Let  $\nu$  be a probability measure on  $\mathbb{R}^n$  and set  $Y$  to be distributed according to  $\nu$ . If  $Z = \|Y\|$  and  $E = (\mathbb{R}^n, \|\cdot\|_{\psi_2})$ , then  $\ell_E \leq c\|Z\|_\infty$ .*

**Proof.** Fix  $\rho$  to be named later and consider the gaussian vector  $G = (g_1, \dots, g_n)$ , where  $(g_i)_{i=1}^n$  are independent, standard gaussian random variables. Let  $\|Z\|_\infty = D$  and since  $\|f\|_{\psi_2} \leq \mathbb{E} \exp(f^2)$  then

$$\frac{\ell_E}{\rho} \leq \mathbb{E}_Y \mathbb{E}_g \exp \left( \frac{\sum_{i=1}^n g_i \langle e_i, Y \rangle}{\rho} \right)^2.$$

Recall that  $\sum_{i=1}^n g_i \langle Y, e_i \rangle$  is distributed as  $g\|Y\|$ , and thus,

$$\begin{aligned} \frac{\ell_E}{\rho} &\leq \mathbb{E}_Y \mathbb{E}_g \left( 1 + \sum_{m=1}^{\infty} \frac{\left( \sum_{i=1}^n g_i \langle Y, e_i \rangle \right)^{2m}}{m! \rho^{2m}} \right) \\ &\leq \left( 1 + \sum_{i=1}^{\infty} \frac{1}{m!} \mathbb{E} |g|^{2m} \left( \frac{D}{\rho} \right)^{2m} \right) \\ &= \mathbb{E} \exp \left( \frac{Dg}{\rho} \right)^2 \leq 2, \end{aligned}$$

if one selects  $\rho = cD$ . Therefore,  $\ell_E \leq cD$ , as claimed.  $\blacksquare$

**Definition 4.8** For two sets  $A, B \subset \mathbb{R}^n$ , let  $N(A, B)$  be the minimal number of translates of  $B$  needed to cover  $A$ , that is, the minimal cardinality of a set  $\{x_1, \dots, x_m\}$  such that  $A \subset \bigcup_{i=1}^m (B + x_i)$ .

Note that if  $B$  is a unit ball of a norm on  $\mathbb{R}^n$  then  $N(A, \varepsilon B)$  are the covering numbers of  $A$  with respect to the metric endowed by  $B$ .

**Corollary 4.9** There exists an absolute constant  $c$  such that for every  $\varepsilon \geq 1/2$ ,

$$\log N(B_2^n, \varepsilon B_E) \leq c \frac{n}{\varepsilon^2},$$

and for  $0 < \varepsilon < 1/2$ ,

$$\log N(B_2^n, \varepsilon B_E) \leq cn \log \left( \frac{1}{\varepsilon} \right),$$

where  $B_E$  is the unit ball of  $(\mathbb{R}^n, \psi_2(\nu))$  and  $B_2^n$  is the Euclidean unit ball.

**Proof.** Recall that if  $Z = \|Y\|$  then  $\|Z\|_\infty \leq c_1 \sqrt{n}$  for a suitable absolute constant. By the dual Sudakov Theorem [18],  $\log N(B_2^n, \varepsilon B_E) \leq c_2 \ell_E / \varepsilon^2$ , and applying Lemma 4.7,  $\ell_E \leq c_3 \sqrt{n}$ , from which the first part of the claim follows. Turning to the second part, by a standard volumetric estimate (see, e.g. [20]), and since  $B_E$  is a unit ball of a norm on  $\mathbb{R}^n$ ,  $N(\frac{1}{2} B_E, \varepsilon B_E) \leq (1/2\varepsilon)^n$ . Therefore,

$$N(B_2^n, \varepsilon B_E) \leq N\left(B_2^n, \frac{1}{2} B_E\right) \cdot N\left(\frac{1}{2} B_E, \varepsilon B_E\right) \leq \exp(c_4 n) \left(\frac{1}{2\varepsilon}\right)^n.$$

$\blacksquare$

Using Corollary 4.9 one can bound  $\gamma_2(F, \psi_2(\nu))$  by applying Theorem 2.7 for the space  $\ell_2^n$  which is 2-convex, and with  $d$  being the  $\psi_2(\nu)$  metric endowed on  $\mathbb{R}^n$ .

**Corollary 4.10** *Let  $\mu$  be an isotropic log-concave measure on  $\mathbb{R}^n$  and set  $\nu$  to be its truncation as above. Then for  $n \leq k \leq \exp(c_1\sqrt{n})$ ,*

$$\gamma_2(S^{n-1}, \psi_2(\nu)) \leq c_2 \sqrt{n \log n},$$

where  $c_1$  and  $c_2$  are absolute constants.

**Proof.** The proof is immediate from Theorem 2.7, the entropy estimate in Corollary 4.9, combined with the fact that for every  $\theta \in S^{n-1}$ ,  $|\langle Y, \theta \rangle| \leq \|Y\| \leq c\sqrt{n}$ , and thus  $\text{diam}(S^{n-1}, \psi_2(\nu)) \leq c\sqrt{n \log k} \leq c\sqrt{n \log n}$ . ■

Let us remark that we believe this estimate is suboptimal by a factor of  $\sqrt{\log n}$ .

Combining Corollary 4.10 with Theorem 3.8, we obtain the following (most likely, suboptimal) estimate of  $\|\mu_k - \mu\|_{F^p}$ , which we only state for  $p > 2$ . This estimate recovers the best known result for  $p > 2$ , and was originally established in [6].

**Theorem 4.11** *For every  $0 < \varepsilon$ ,  $0 < \delta < 1$  and  $p > 2$  there exists a constant  $c(\varepsilon, \delta, p)$  for which the following holds. With probability at least  $1 - \delta$ , if  $k \geq k_0$ ,*

$$\sup_{\theta \in S^{n-1}} \left| \frac{1}{k} \sum_{i=1}^k |\langle X_i, \theta \rangle|^p - \mathbb{E} |\langle X, \theta \rangle|^p \right| < \varepsilon,$$

provided that  $k_0 \geq c(\varepsilon, \delta, p)n^{p/2} \log n$  for  $p > 2$ .

**Proof.** Let  $n \leq k \leq \exp(c_1\sqrt{n})$ . Using the notation of Theorem 3.8, observe that  $\gamma_2(S^{n-1}, \psi_2(\nu)) \leq c_2 \sqrt{n \log n}$ ,  $\text{diam}(S^{n-1}, \psi_1) \leq c_2$ ,  $H_k \leq c_2 \sqrt{n}$ . Also, if  $k \geq c_3 n \log n$ , then for  $p > 2$ ,  $\theta$  can be taken as  $\theta \leq c_4 \log \log n$ , from which the claim is evident. ■

Let us remark that if one could select  $k \leq cn \log n$  it would be possible to take  $\theta$  at the level of an absolute constant. This would be the case if the logarithmic term in the estimate on  $\gamma_2(S^{n-1}, \psi_2(\nu))$  were to be removed and would lead to the optimal bound for any  $p > 1$ .

## References

- [1] F. Barthe, O. Guédon, S. Mendelson, A. Naor, A probabilistic approach to the geometry of the  $\ell_p^n$  ball, *Ann. Probab.* 33(2), 480-513, 2005.
- [2] C. Borel, The Brunn-Minkowski inequality in Gauss space, *Invent. Math.* 30 207-216, 1975.
- [3] J. Bourgain, Random points in isotropic convex bodies, in *Convex Geometric Analysis* (Berkeley, CA, 1996) *Math. Sci. Res. Inst. Publ.* 34 (1999), 53-58.
- [4] A.A. Giannopoulos, V.D. Milman, Concentration property on probability spaces, *Adv. Math.* 156, 77-106, 2000.
- [5] E. Giné and J. Zinn, Some limit theorems for empirical processes, *Ann. Probab.* 12(4), 929-989, 1984.
- [6] O. Guédon, M. Rudelson,  $L_p$  moments of random vectors via majorizing measures, preprint.
- [7] R. Kannan, L. Lovász, M. Simonovits, Random walks and  $O^*(n^5)$  volume algorithm for convex bodies, *Random structures and algorithms*, 2(1) 1-50, 1997.
- [8] M. Ledoux: *The concentration of measure phenomenon*, Mathematical Surveys and Monographs, Vol 89, AMS, 2001.
- [9] M. Ledoux, M. Talagrand, *Probability in Banach spaces. Isoperimetry and processes*, *Ergebnisse der Mathematik und ihrer Grenzgebiete (3)*, vol. 23. Springer-Verlag, Berlin, 1991.
- [10] F. Lust-Piquard, G. Pisier, Non-commutative Khinchine and Paley inequalities, *Ark. Mat.* 29, 241-260, 1991.
- [11] S. Mendelson, A. Pajor, M. Rudelson, On the Geometry of random  $\{-1, 1\}$ -polytopes, *Discrete and Computational Geometry*, 33(3) 365-379, 2005.
- [12] S. Mendelson, A. Pajor, N. Tomczak-Jaegermann, Reconstruction and subgaussian operators, preprint.
- [13] V. Milman, Spectrum of a position of a convex body and linear duality relations, *Israel Mathematical Conference Proceedings* Vol 3, 151-161, 1990.
- [14] V. Milman, A. Pajor, Regularization of star bodies by random hyperplane cut off, *Studia Math.* 159, 247-261, 2003.
- [15] V. Milman, G. Pisier, Banach spaces with a weak cotype 2 property, *Israel J. Math* 54(2), 139-158, 1986.
- [16] V.D. Milman, G. Schechtman, *Asymptotic theory of finite dimensional normed spaces*, Lecture Notes in Mathematics 1200, Springer, 1986.
- [17] A. Pajor, N. Tomczak-Jaegermann, Subspaces of small codimension of finite-dimensional Banach spaces, *Pro. AMS.* 97(4), 637-642, 1986.
- [18] A. Pajor, N. Tomczak-Jaegermann, Remarques sur les nombres d'entropie d'un opérateur et de son transposé, *C. R. Acad. Sci Paris* 301, 743-746, 1985.
- [19] G. Paouris, private communication.
- [20] G. Pisier, *The volume of convex bodies and Banach space geometry*, Cambridge University Press, 1989.

- [21] M. Rudelson, Random vectors in the isotropic position, *J. Funct. Anal.* 164, 60-72, 1999.
- [22] M. Talagrand, Sharper bounds for Gaussian and empirical processes, *Ann. Probab.* 22(1), 28-76, 1994.
- [23] M. Talagrand, Majorizing measures: the generic chaining, *Ann Probab.* 24(3) 1049-1103, 1996.
- [24] M. Talagrand, *The generic chaining*, Springer, 2005.
- [25] A.W. Van der Vaart, J.A. Wellner, *Weak convergence and empirical processes*, Springer Verlag, 1996.